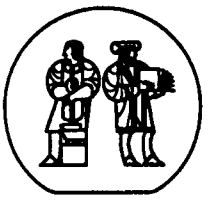

VLSI Memo No. 89-562
September 1989

Parallel Distributed Networks for Image Smoothing and Segmentation in Analog VLSI

A. Lumsdaine, J. Wyatt, and I. Elfadel

Abstract

Image smoothing and segmentation algorithms are frequently formulated as optimization problems. Linear and nonlinear (reciprocal) *resistive networks* have solutions characterized by an extremum principle. Thus, appropriately designed networks can *automatically* solve certain smoothing and segmentation problems in robot vision. This paper considers switched linear resistive networks and nonlinear resistive networks for such tasks. A new derivation of the latter network type from the former is given via an intermediate stochastic formulation, and a new result relating the solution sets of the two is given for the "zero temperature" limit. We then present simulation studies of several continuation methods that can be gracefully implemented in analog VLSI and that seem to give "good" results for these non-convex optimization problems.

Microsystems Technology Laboratories | Massachusetts Institute of Technology | Cambridge Massachusetts 02139 | Room 39-321 Telephone (617) 253-0292

Acknowledgements

This work was supported by Defense Advanced Research Projects Agency Contract N00014-87-K-825, National Science Foundation Grant MIP-88-14612, and E. I. DuPont de Nemours and Co. A. Lumsdaine was also supported by an AEA/Dynatech faculty development fellowship.

Author Information

Elfadel: Department of Electrical Engineering and Computer Science, Room 36-863, MIT, Cambridge, MA 02139. (617) 253-2631.

Lumsdaine: Department of Electrical Engineering and Computer Science, Room 36-893, MIT, Cambridge, MA 02139. (617) 253-7307.

Wyatt: Department of Electrical Engineering and Computer Science, Room 36-864, MIT, Cambridge, MA 02139. (617) 253-6718.

To appear in Proceedings
of the 28th IEEE Conference
on Decision and Control, Tampa, FL,
December 13-15, 1989.

Parallel Distributed Networks for Image Smoothing and Segmentation in Analog VLSI

A. Lumsdaine J. Wyatt I. Elfadel

Research Laboratory of Electronics
Dept. of Electrical Engineering and Computer Science
Massachusetts Institute of Technology
Cambridge, MA 02139

ABSTRACT

Image smoothing and segmentation algorithms are frequently formulated as optimization problems. Linear and nonlinear (reciprocal) *resistive networks* have solutions characterized by an extremum principle. Thus, appropriately designed networks can *automatically* solve certain smoothing and segmentation problems in robot vision. This paper considers switched linear resistive networks and nonlinear resistive networks for such tasks. A new derivation of the latter network type from the former is given via an intermediate stochastic formulation, and a new result relating the solution sets of the two is given for the "zero temperature" limit. We then present simulation studies of several continuation methods that can be gracefully implemented in analog VLSI and that seem to give "good" results for these non-convex optimization problems.

1 Introduction

One of the most important, yet most difficult, early vision tasks is that of image smoothing and segmentation. Smoothing is necessary to remove noise from an input image so that reliable processing in subsequent stages is facilitated. However, indiscriminate smoothing will blur the entire image, including edges (corresponding to object boundaries) which are necessary for later stages of processing. Many researchers are currently seeking to develop algorithms that smooth in a piecewise manner, respecting edges. There are two main approaches taken — stochastic, [Geman & Geman], [Marroquin], [Cohen], and deterministic [Blake & Zisserman], [Blake], [Perona & Malik]. The former relies on such methods as simulated annealing to accomplish the minimization. The deterministic approach, on the other hand, relies on the application of continuation methods [Ortega & Rheinboldt] to certain nonlinear systems, or in the case of [Koch *et al*], on using a neural network similar to that of Hopfield and Tank [Tank & Hopfield].

Although efficient computation techniques exist for numerically computing the solutions to vision problems [Terzopoulos], even the fastest algorithms running on a parallel supercomputer (such as the Connection Machine [Hillis]) do not approach real-time performance. The motivation of this work is to produce solutions to the smoothing and segmentation problem that are amenable to analog VLSI network implementation [Koch *et al*], [Poggio & Koch], [Horn].

The outline of the paper is as follows. In Section 2, we present the smoothing and segmentation task as a minimization problem. In Section 3, we present methods for solving the minimization problem and discuss network implementations of these methods. Simulation results are provided in Section 4 and the conclusion is given in Section 5.

2 Minimization

The difficulty with using a *linear* network for image smoothing is that noise and signal are equally smoothed so that edges become blurred. We therefore seek a method for segmenting the signal into regions which can be smoothed separately. One technique for doing this is to introduce a line process (i.e., a set of binary variables) which selectively breaks the smoothness constraint at given locations. This method appears widely in the literature, e.g., [Geman & Geman], [Koch *et al*], [Marroquin], and [Marroquin *et al*].

For simplicity of notation, all equations in this paper are formulated for the one-dimensional case. The results generalize trivially to two dimensions, and the simulation results are for the two-dimensional case.

The smoothing and segmentation problem with the line process can be treated as a minimization problem. Let $\mathbf{u} \in \Re^N$ be the input image, $\mathbf{y} \in \Re^N$ be the output image, and $\mathbf{l} \in \Re^{N-1}$ be the line process, where the line process variable l_i assumes the values $\{0, 1\}$ depending on whether the smoothness penalty between nodes i and $i+1$ is enforced or not. Consider the following cost function:

$$J(\mathbf{y}, \mathbf{l}|\mathbf{u}) = \frac{1}{2}[F(\mathbf{y}|\mathbf{u}) + S(\mathbf{y}, \mathbf{l}) + L(\mathbf{l})] \tag{1}$$

where F, S, and L are the "fidelity," "smoothness," and "line" penalty terms, respectively, i.e.,

$$F(\mathbf{y}|\mathbf{u}) = \lambda_f \sum_{i=1}^{N} (y_i - u_i)^2 \tag{2}$$

$$S(\mathbf{y}, \mathbf{l}) = \lambda_s \sum_{i=1}^{N-1} (y_i - y_{i+1})^2 (1 - l_i) \tag{3}$$

$$L(\mathbf{l}) = \lambda_l \sum_{i=1}^{N-1} l_i. \tag{4}$$

The expression (1) can be minimized with respect to $\mathbf{y}$ for fixed $\mathbf{l}$ by differentiating with respect to each y_i and setting the derivatives to zero. This produces the following system of equations:

$$\lambda_f(y_i - u_i) + \lambda_s(y_i - y_{i-1})(1 - l_{i-1}) + \lambda_s(y_i - y_{i+1})(1 - l_i) = 0, \tag{5}$$

with appropriate modifications at the boundaries $i = 1$ and $i = N$. Notice that (5) can be viewed as the KCL relation at

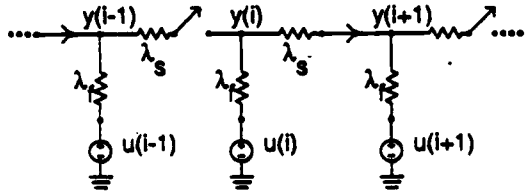

Figure 1: A simple smoothing network with switches. The vertical and horizontal resistors have conductances λ_f and λ_s, respectively.

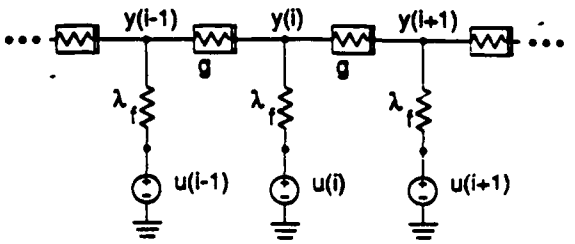

Figure 2: Nonlinear network topology.

every node of a resistive ladder in which the horizontal resistive elements have switches (corresponding to a line process element) associated with them. A network for computing y given l is shown in Figure 1; similar networks have appeared in [Koch *et al*] and [Marroquin *et al*]. This type of network will be referred to as a resistor-with-switch (or RWS) network. For any setting of the switches, the network will minimize the cost function with respect to y. The difficulty is in minimizing with respect to l, because the network solution can be any one of a number of local minima.

Much has been said in the literature in regard to finding a global minimum to (1) by stochastic and deterministic methods. These techniques are necessary to find the minimizing l — minimizing with respect to y *given* l only requires the solution of a linear system. The deterministic approaches rely on the fact that the minimization problem can be recast into one in which the line process variables have been eliminated. The latter will be studied here since they appear to lead to practical VLSI implementations.

2.1 Resistive Fuse Elements

The line process variables can be removed from (1) by straightforward algebraic manipulations. In fact, Blake and Zisserman demonstrated that the original cost function $J(\mathbf{y},\mathbf{l}|\mathbf{u})$ containing real and boolean variables is intimately related to the following cost function containing only real variables:

$$K(\mathbf{y}|\mathbf{u}) \triangleq \frac{1}{2}\left[\lambda_f \sum_{i=1}^{N}(y_i-u_i)^2 + \sum_{i=1}^{N-1} G(y_i-y_{i+1})\right], \quad (6)$$

where

$$G(v) = \begin{cases} \lambda_s v^2, & |v| < \sqrt{\frac{\lambda_f}{\lambda_s}} \\ \lambda_f, & \text{otherwise} \end{cases}. \quad (7)$$

The line process is found *a posteriori* according to:

$$l_i = \begin{cases} 0, & |y_i - y_{i+1}| < \sqrt{\frac{\lambda_f}{\lambda_s}} \\ 1, & \text{otherwise} \end{cases}. \quad (8)$$

Note that K is a *non-convex* cost function with respect to y.

Apart from instances in which solutions occur at points where G is not differentiable, the minimum of K is to be found among those points where $\nabla K(\mathbf{y}|\mathbf{u}) = 0$, i.e.,

$$\lambda_f(y_i - u_i) + g(y_i - y_{i-1}) + g(y_i - y_{i+1}) = 0, \quad (9)$$

where $g(v) = \frac{1}{2}\frac{d}{dv}G(v)$.

Equation (9) can also be viewed as the KCL relation at each node of a nonlinear resistive network with the topology

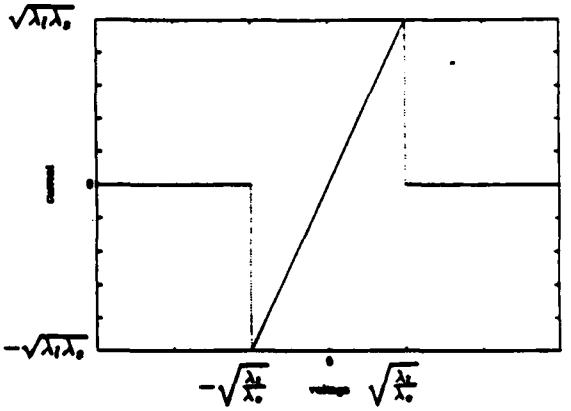

Figure 3: Characteristic of the discontinuous nonlinear resistor known as a "resistive fuse." The dotted vertical lines are *not* part of the constitutive relation.

illustrated in Figure 2. The nonlinear resistor characteristic, $g(v)$, is that of a linear resistor that becomes an open circuit when the voltage across it exceeds a certain threshold, as shown in Figure 3. Then, in electrical terms, G is twice the *co-content* function for this nonlinear resistor [Millar], i.e., $G(v) = 2\int_0^v g(u)\,du$. An element with this type of behavior is sometimes called a *resistive fuse* and has actually been implemented in analog VLSI [Harris *et al*]. A network incorporating resistive fuses will be referred to as an RWF network, i.e., a *resistor-with-fuse* network.

For a given cost function, one can construct corresponding RWS and RWF networks. For every solution of an RWF network, there exists a similar solution to the corresponding RWS network, but there are switch configurations of an RWS network for which there is *no* corresponding solution in a corresponding RWF network. The question then arises whether restricting attention to the RWF network might cause one to overlook a solution to the RWS network that is in fact the global minimum. The answer is no, by the following proposition:

Proposition 1 Consider a cost function $J(\mathbf{y},\mathbf{l}|\mathbf{u})$ as specified in (1) and the corresponding RWS and RWF networks specified by (5) and (9). If a solution $\mathbf{y}^*$ exists for the RWS network which is not also a solution to the RWF network, then $\mathbf{y}^*$ is *not* a local minimum to J, meaning that changing the setting of a *single* (appropriately selected) switch in the RWS network[6] will produce a new solution (with a new value of y) for which the value of J is strictly lower.

In order to complete the proof, we need the following lemma:

Lemma 1 (Local Measurement Principle) Consider a one-dimensional or two-dimensional network of the type shown in Figure 1, in which an arbitrary number of switches (≥ 1) are open. Let us consider the change in the *power dissipated* in the network

$$\begin{aligned} P &= \lambda_f \sum_{i=1}^{N} (y_i - u_i)^2 + \lambda_s \sum_{i=1}^{N-1} (y_i - y_{i+1})^2 (1 - l_i) \\ &= 2(F + S) \end{aligned} \tag{10}$$

caused by closing a single switch. Let P^- be the value of P with the switch open, P^+ be the value of P with the switch closed, and define $\Delta P = P^+ - P^-$. Let i_{sc} be the current through the switch when it is closed and let v_{oc} be the voltage across the switch when it is opened (after the network has settled). Then the increase in dissipation which results from closing the switch is

$$\Delta P = v_{oc} i_{sc} > 0. \tag{11}$$

Remark: This is a startling result. The local measurement principle implies that one can measure the *global* change in the network cost function due to a switch change merely by taking two measurements *at the switch*. Both proofs below use circuit theory techniques, but can also be carried out, albeit laboriously, by mathematical arguments divorced from a network realization, e.g., the proof of Lemma 1 via a rank 1 perturbation method in [Elfadel].

Proof of Lemma 1: Define v_k^- and i_k^- to be the network branch voltages and currents when the switch is open. Define v_k and i_k to be the network branch voltages and branch currents when the switch is closed. Define $\Delta v_k = v_k - v_k^-$ and $\Delta i_k = i_k - i_k^-$. By Tellegen's theorem [Penfield *et al*],

$$\sum_{\substack{\text{all} \\ \text{branches}}} [v_k \Delta i_k - i_k \Delta v_k] = 0. \tag{12}$$

Group the terms in (12) according to branch element, and note that

$$\begin{aligned} &\sum_{\substack{\text{voltage} \\ \text{sources}}} [v_k \Delta i_k - i_k \Delta v_k] + \sum_{\text{resistors}} [v_k \Delta i_k - i_k \Delta v_k] \\ &+ \sum_{\substack{\text{fixed} \\ \text{switches}}} [v_k \Delta i_k - i_k \Delta v_k] + v_{sw} \Delta i_{sw} - i_{sw} \Delta v_{sw} = 0, \end{aligned} \tag{13}$$

where the subscript "sw" refers to the switch that is being closed and "fixed switches" to all others. To simplify (13), note that $\Delta v_k = 0$ for the voltage sources, v_k and Δv_k vanish for closed switches, i_k and Δi_k vanish for open switches, and for the resistors:

$$v_k \Delta i_k - i_k \Delta v_k = R_k i_k \Delta i_k - i_k R_k \Delta i_k = 0. \tag{14}$$

Equation (13) then becomes

$$\begin{aligned} 0 &= \sum_{\substack{\text{voltage} \\ \text{sources}}} [v_k \Delta i_k] + v_{sw} \Delta i_{sw} - i_{sw} \Delta v_{sw} \\ &= \sum_{\substack{\text{voltage} \\ \text{sources}}} [v_k \Delta i_k] + v_{sw} (i_{sw} - i_{sw}^-) - i_{sw} (v_{sw} - v_{sw}^-) \\ &= \sum_{\substack{\text{voltage} \\ \text{sources}}} [v_k \Delta i_k] + v_{sw}^- i_{sw} \end{aligned} \tag{15}$$

The summation term in (15) is just the change in power delivered to the network, i.e., $-\Delta P$, and $v_{sw}^- i_{sw} = v_{oc} i_{sc}$. Therefore,

$$\Delta P = v_{oc} i_{sc}. \tag{16}$$

∎

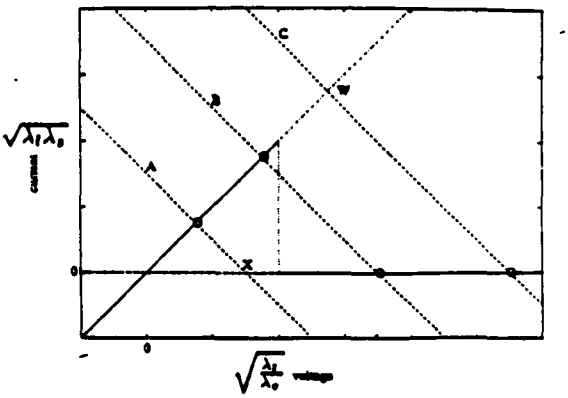

Figure 4: Load-Line diagram for RWS network and RWS network with resistive fuse substituted for one resistor-switch composite element. The dashed lines A, B, and C are possible load-lines representing the behavior of an RWS network as seen by one resistor-switch pair. The six marked points indicate possible solutions, depending on switch position. If a resistive fuse element (with characteristic $i = g(v)$, shown with a solid line) is substituted for the resistor-switch pair, the four circled solutions remain, while the solutions marked X and W disappear.

Proof of Proposition 1: Consider any RWS network with any input u, switch configuration l, and corresponding network solution $\mathbf{y}^*$, such that $\mathbf{y}^*$ is not a solution of the corresponding RWF network. Then there must exist *some* resistor-switch composite element (element q, say), such that $\mathbf{y}^*$ is no longer a network solution if a resistive fuse is substituted in its place. Make such a substitution and then consider the load-line describing the remainder of the linear RWS network as seen from this location. The two possible cases marked in Figure 4 are Case X, in which switch q was open in the original RWS network, and Case W, in which switch q was closed. Note that the area in the first quadrant under the triangle is $\frac{1}{2}\lambda_l$. In Case X, closing the switch in the original RWS network would have caused the solution to move to the circled point on line A. By Lemma 1 the change would be

$$J_{closed} - J_{open} = \frac{1}{2}[v_{oc} i_{sc} - \lambda_l] < 0, \tag{17}$$

where the inequality follows from the fact that $\frac{1}{2} v_{oc} i_{sc}$ (the area under the line connecting the origin to (v_{oc}, i_{sc})) is less than $\frac{1}{2}\lambda_l$ (the area under the triangle). For Case W, similar reasoning shows $J_{open} - J_{closed} < 0$ if opening the switch causes the network solution to move from point W to the circled point on line C. Thus points X and W in Figure 4 are not local minima of J. ∎

Remark: The converse of the proposition is not true. If the network solution lies on load-line B, one intersection point or the other will generally have lower cost for the RWS network, yet both are valid solutions to the RWF network.

2.2 Marginal Distribution of Reconstructed Intensities in a Stochastic Formulation

There is an alternate method for deriving the RWF network from the RWS network: calculating the marginal distribution of reconstructed intensities in a stochastic formulation. In this

approach, one postulates that the *a posteriori* joint probability distribution for y and l given u is of the Gibbs form

$$p(\mathbf{y},\mathbf{l}) = c_1 e^{-\beta[F(\mathbf{y}|\mathbf{u})+S(\mathbf{y},\mathbf{l})+L(\mathbf{l})]} \tag{18}$$

where c_1 is a normalizing constant, and F, S, and L are given in (2) – (4). The marginal *a posteriori* distribution of y is obtained in the usual way:

$$\begin{aligned} p(\mathbf{y}) &= \sum_{\mathbf{l}\in C} p(\mathbf{y},\mathbf{l}) \\ &= c_1 \sum_{\mathbf{l}\in C} e^{-\beta[F(\mathbf{y}|\mathbf{u})+S(\mathbf{y},\mathbf{l})+L(\mathbf{l})]} \end{aligned} \tag{19}$$

where C is the set of all corners of the unit $(N-1)$-dimensional hypercube.

A useful closed form expression for $p(\mathbf{y})$ appears below:

Lemma 2

$$p(\mathbf{y}) = c_2 e^{-\beta[F(\mathbf{y}|\mathbf{u})+J_2(\mathbf{y})]} \tag{20}$$

where c_2 is a normalizing constant, $F(\mathbf{y}|\mathbf{u})$ is given in (2), and

$$J_2(\mathbf{y}) = \frac{1}{\beta}\sum_{i=1}^{N-1} \ln\left(\frac{1+e^{\beta\lambda_l}}{1+e^{\beta[\lambda_l-\lambda_s(y_i-y_{i+1})^2]}}\right). \tag{21}$$

The proof of the lemma requires the following fact, which can easily be verified.

Fact: Let $\mathbf{b} = (b_1,\ldots,b_n)$ be a vector of n binary variables, $b_i \in \{0,1\}$, and let C be the set of all such vectors. Then for any $\mathbf{r} \in \Re^n$,

$$\sum_{\mathbf{b}\in C} e^{\mathbf{b}\cdot\mathbf{r}} = \prod_{i=1}^{n}(1+e^{r_i}), \tag{22}$$

where $\mathbf{b}\cdot\mathbf{r}$ is the standard inner product.

Proof of Lemma 2: (It's only algebra)

The terms being summed in (19) can be decomposed as follows:

$$\begin{aligned} & e^{-\beta[\lambda_f\sum_{i=1}^{N}(y_i-u_i)^2+\lambda_s\sum_{i=1}^{N-1}(y_i-y_{i+1})^2]} \\ & \times \left\{e^{-\beta\sum_{i=1}^{N-1} l_i[\lambda_l-\lambda_s(y_i-y_{i+1})^2]}\right\}. \end{aligned} \tag{23}$$

Using 22, the term in braces sums to

$$\exp\left(-\beta\{-\frac{1}{\beta}\sum_{i=1}^{N-1}\ln(1+e^{-\beta[\lambda_l-\lambda_s(y_i-y_{i+1})^2]})\}\right) \tag{24}$$

and further algebraic manipulation shows that

$$p(\mathbf{y}) = c_1 e^{-\beta[F(\mathbf{y}|\mathbf{u})+\tilde{J}_2(\mathbf{y})]}, \tag{25}$$

where

$$\tilde{J}_2(\mathbf{y}) = \frac{1}{\beta}\sum_{i=1}^{N-1}\ln\left(\frac{1}{1+e^{\beta[\lambda_l-\lambda_s(y_i-y_{i+1})^2]}}\right)+(N-1)\lambda_l. \tag{26}$$

Absorbing an additive term into the normalizing constant c_2, the lemma was stated in (21) in terms of

$$J_2(\mathbf{y}) \triangleq \tilde{J}_2(\mathbf{y}) + \frac{N-1}{\beta}\left[\ln(1+e^{\beta\lambda_l})-\beta\lambda_l\right], \tag{27}$$

which is constructed so that $J_2(\mathbf{0}) = 0$. This is a necessary step if we are to later interpret $J_2(\mathbf{y})$ as the co-content function of a set of nonlinear resistors. ∎

Remark: Equation (20) suggests a new cost function

$$K_\beta(\mathbf{y}|\mathbf{u}) = \frac{1}{2}\left[F(\mathbf{y}|\mathbf{u})+J_2(\mathbf{y})\right]. \tag{28}$$

The minima of $K_\beta(\mathbf{y}|\mathbf{u})$ are obtained from the set of points satisfying

$$\nabla K_\beta(\mathbf{y}|\mathbf{u}) = 0. \tag{29}$$

Taking the i-th component of (29) gives:

$$\lambda_f(y_i-u_i)+g_\beta(y_i-y_{i-1})+g_\beta(y_i-y_{i+1}) = 0, \tag{30}$$

where

$$g_\beta(v) = \frac{\lambda_s v}{1+e^{-\beta[\lambda_l-\lambda_s v^2]}}. \tag{31}$$

Equation (30) can be considered the KCL relation at each node of a nonlinear network having vertical linear resistive elements with conductance λ_f and horizontal nonlinear elements with constitutive relation $i = g_\beta(v)$. In this case, $K_\beta(\mathbf{y}|\mathbf{u})$ is the *total co-content* of the network. Notice that as $\beta \to \infty$, we recover the RWF network, i.e., $K_\infty(\mathbf{y}|\mathbf{u}) = K(\mathbf{y}|\mathbf{u})$. Moreover, we have defined a family of β-dependent resistive elements, illustrated in Figure 6, that can be used in continuation methods.

The expressions (26) and (31) have also been obtained from the RWS network by a somewhat different stochastic method involving repeated use of the mean-field approximation [Geiger & Girosi]. No approximations are involved in the marginal density formulation above.

3 Solution Methods and Network Implementations

The resistive fuse and marginal distribution approaches produced switch-free nonlinear networks with identical topologies (see Figure 2) but with different constitutive relations for the nonlinear elements. For either network, multiple solutions generally exist. On the theoretical side this is a difficulty because we are trying to find the global minimum of a specific cost function. On the practical side this is a difficulty because the solution that is obtained by a physical network realization will depend strongly on such things as parasitic capacitances and other characteristics of the network over which we have little control. We therefore seek some modification of the network that will allow us to exercise some control over the solution it finds. In this section, we apply continuation methods to the nonlinear smoothing and segmentation networks.

3.1 Example — A Special Case

The simplest special case that nonetheless provides insight into the phenomenon of multiple solutions is the response of a one-dimensional network to a step edge input, i.e.,

$$u_i = \begin{cases} u_H & i \le k \\ u_L & i > k \end{cases}, \tag{32}$$

for some $k < N$. This corresponds to a step of $u_H - u_L > 0$ between nodes k and $k+1$ and serves as a model for the simplest two-dimensional edge, i.e., a step that extends across the entire network and is parallel to one of the network "axes".

For the step input described above, the one-dimensional network has a simple circuit equivalent, shown in Figure 5. The simplification proceeds as follows. First, we *assume* that the signal is "well-smoothed" on either side of the step so that

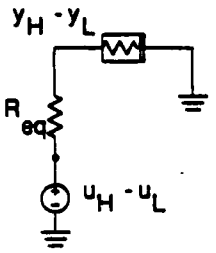

Figure 5: Thevenin equivalent circuit for nonlinear smoothing and segmentation network with step input.

each nonlinear element can be replaced by an equivalent linear resistance whose value is the incremental resistance of the nonlinear element about zero volts. The network elements on either side of the step are then replaced by their Thevenin equivalents, which are combined into a single linear element and voltage source. The simplified network will be referred to as the zero-dimensional case. Analysis of the behavior of the network to a step input is reduced to solving the KCL equation at one node: some insight into the circuit behavior can be gained by using load-line techniques (see Figure 4 – Figure 7).

This "linear load-line assumption" holds *exactly* only for the RWS network with fixed switch positions and for the marginal distribution network with $\beta = 0$. For the RWF network and for the marginal distribution network with $\beta \to \infty$, it is exact over the limited voltage range in which no new discontinuities are introduced into y. Otherwise, it is only an approximation and its applicability to other cases of interest must be individually determined.

3.2 Continuation Methods

We seek a modification to the networks so that the solution will be repeatable and also be visually and quantitatively "good." One technique that works well within the context of smoothing and segmentation is to apply a continuation method to the network [Ortega & Rheinboldt].

A continuation (sometimes called "deterministic annealing") can be realized in network form by the simultaneous application of a given homotopy (continuous deformation) to some or all of the circuit elements. Two types of continuations are particularly appropriate for our class of nonlinear networks. Assume we have a network with horizontal nonlinear resistors whose constitutive relation is described by $i = g(v)$, and vertical linear resistors with conductance λ_f. Consider the following two homotopies for the horizontal and vertical elements, respectively:

CH: Replace $g(v)$ with $g^{(p)}(v)$, $p \in [a, b]$, such that $g^{(a)}$ constrains the network to have a unique solution and that $g^{(b)}(v) = g(v)$;

CV: Replace λ_f with $\lambda_f^{(p)}$, $p \in [a, b]$, such that $\lambda_f^{(a)}$ constrains the network to have a unique solution and that $\lambda_f^{(b)} = \lambda_f$.

Note that CH and CV define *where* the homotopies are applied in the network to produce a continuation; we are still free to decide the specific form of the homotopy.

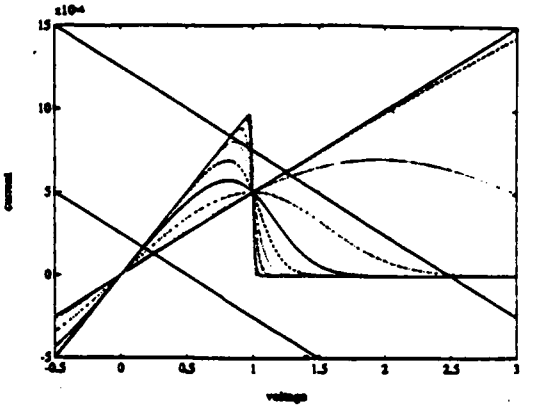

Figure 6: Approximate load-line plot for marginal distribution network with β-Continuation. The lines with negative slope represent the load-lines for two different input values ($0.5\,V$ and $2.5\,V$). The nonlinear resistor is shown for various values of β. For $\beta = 0$, the nonlinear resistor acts as a linear resistor. As $\beta \to \infty$, the nonlinear resistor characteristic becomes that of the RWF.

3.3 β-Continuation

Blake and Zisserman suggest a CH continuation method — the so-called "graduated non-convexity" algorithm, or GNC. There are several apparent weaknesses to using the GNC algorithm in network form, however. First, there is no reason to expect that the specific continuation used by GNC will produce the global minimum or that it will even produce a "good" minimum. Second, and more importantly, the initial state (i.e., the initial value of the continuation parameter) of the network realization of GNC is not guaranteed to have a unique solution.

On the other hand, the marginal distribution derivation of our nonlinear network provides a natural homotopy for realization of the CH continuation. For $\beta = 0$, the network with elements described by (31) is linear, whereas for $\beta = \infty$, the elements become identical to those in Figure 3 and will (locally) solve our minimization problem. This suggests using β directly as the continuation parameter for a CH continuation for solving (30) and hence (9). Furthermore, because of the way this continuation was derived, one might expect that it would do a good job of seeking the global cost minimum.

Some insight into the behavior of this type of network can be gained by examining the zero-dimensional case. Figure 6 shows the marginal distribution nonlinear resistor characteristic for various values of β, along with two load-lines representing two different values of the input. As β is taken from 0 to ∞, the solution will follow the continuous path represented by the intersection of the resistor curve and the load-line. In this example, the smaller step will be smoothed, and the larger step will be segmented.

3.4 λ_f-Continuation

The CV continuation can be realized in a straightforward manner by performing a homotopy on the vertical resistors in the network. In particular, we begin with the resistors having infinite (or sufficiently large) conductance so that the network has only one solution, namely $y = u$ (or $y \approx u$). Then, we continuously decrease the value of the conductance to λ_f.

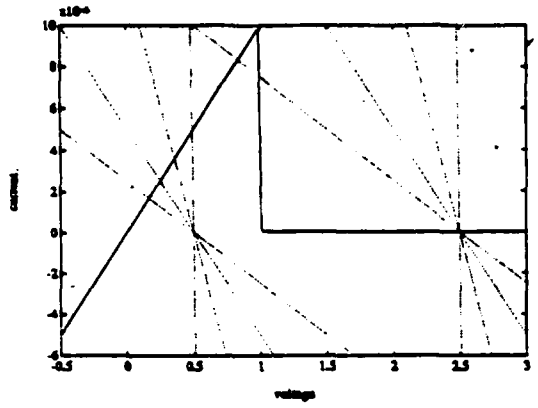

Figure 7: Approximate load-line plot for λ_f-continuation. Two sets of load-lines are shown, each set for a different value of the input (the load-lines intersect the $g(v) = 0$ line at the value of the input voltage — $0.5\,V$ and $2.5\,V$). As λ_f is decreased, the load-lines rotate counter-clockwise.

Examination of the zero-dimensional case provides some insight into the behavior of this type of network. Figure 7 shows the marginal distribution nonlinear resistor characteristic for large β, along with two sequences of load-lines representing two different values of the input. As $\lambda_f^{(p)}$ is taken from $\lambda_f^{(a)} = \lambda_0$ to $\lambda_f^{(b)} = \lambda_f$, the solution will follow the continuous path represented by the intersection of the resistor curve and the load-line. In this example, the smaller step will be smoothed, and the larger step will be segmented.

3.5 Dynamic Solution

There is one final method for minimizing (28) and realizing (30) as a network: embed the static network into a dynamic one so that (28) is a Liapunov function of the dynamic network. We do this by adding a fixed value linear capacitor to ground at each node of the network (see Figure 8). This gives

$$C\frac{d}{dt}y_i = \lambda_f(y_i - u_i) + g_\beta(y_i - y_{i-1}) + g_\beta(y_i - y_{i+1}) \quad (33)$$

as the KCL relationship at each node of the network. At the final state of the network, $\frac{d}{dt}y = 0$, which is precisely what we want. We still need to specify the initial condition and there are two that are interesting, namely $y(0) = 0$ and $y(0) = u$. It can easily be shown that the network always settles to equilibrium for any initial condition and for any fixed set of parameters.

Note that for values of $\beta < \infty$, and for $y(0) = u$, this network is essentially a modified form of the one presented in [Perona & Malik]. This network has an advantage over that presented in [Perona & Malik], however: it settles to a nontrivial equilibrium point that (at least locally) minimizes a specified cost function.

4 Numerical Experiments

Our goals in this section are two-fold: to measure how well the networks actually minimize the cost function given in (1) and to qualitatively observe how the networks smooth and segment images.

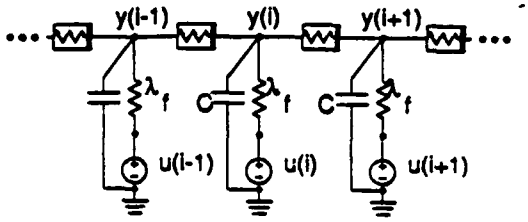

Figure 8: A dynamic smoothing network.

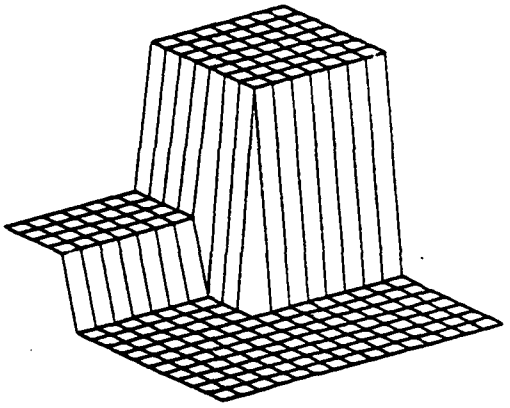

Figure 9: Input image used for experiments.

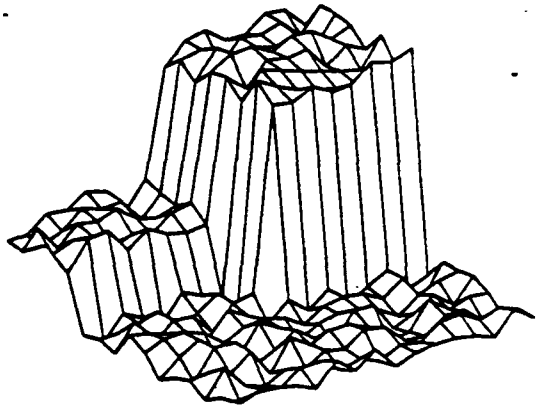

Figure 10: Input image with additive noise.

The experiments were all conducted on 16 x 16 two-dimensional grids. Figure 9 shows the synthetic image used for the experiments. The small step is $1\,V$ in height and the large step is $3\,V$. The original image was then corrupted by the addition of $0.5\,V$ of uniformly distributed noise and is shown in Figure 10. The noisy signal was used as input for all experiments.

The results of three experiments are included here. For each experiment, a cost function was determined, and the β-continuation, λ_f-continuation, and dynamic networks were constructed according to the cost function. Then, the networks were each simulated using the input image shown in Figure 10 (for the dynamic network, the initial conditions $y(0) = 0$ and $y(0) = u$ were both tested). In order to es-

Experiment	Parameter λ_f	λ_s	λ_l
Expt 1	1.0×10^{-3}	1.0×10^{-3}	1.0×10^{-3}
Expt 2	1.0×10^{-3}	1.0×10^{-3}	2.5×10^{-3}
Expt 3	3.0×10^{-3}	1.0×10^{-3}	5.0×10^{-4}

Table 1: Parameter values for three experiments.

Network	Cost Expt 1	Expt 2	Expt 3
Anneal	1.187×10^{-2}	2.259×10^{-2}	4.129×10^{-3}
β-Cont	1.187×10^{-2}	2.259×10^{-2}	4.129×10^{-3}
λ_f-Cont	1.187×10^{-2}	2.387×10^{-2}	6.084×10^{-3}
Dyn, $y(0)=0$	1.187×10^{-2}	2.906×10^{-2}	4.129×10^{-3}
Dyn, $y(0)=u$	1.187×10^{-2}	2.387×10^{-2}	6.084×10^{-3}

Table 2: Values of cost functions for the three experiments. The global cost minimum (Anneal) is also included.

tablish a baseline result, a solution we believe to be the global minimum solution of the RWS was calculated with a simulated annealing algorithm. Table 1 shows the network parameters used for the three experiments. The value of β was fixed at 2×10^4 for all experiments except the β-continuation, in which it was increased from 0 to 2×10^4.

Solutions obtained by the different nonlinear networks can be compared as follows:

1. Given a cost function, construct the corresponding nonlinear networks, and in addition, construct a *corresponding* RWS network;
2. Provide each network with the same input and allow each network to attain its solution;
3. For each nonlinear network, transfer the line process solution obtained to the RWS network by setting the switches accordingly;
4. Allow the RWS to attain its voltage solution and compute the resulting cost — it is this cost that is used for purposes of comparison.

By doing this, we measure how well the network solves the original problem (1), and in addition are provided with a uniform method for comparing the solutions obtained by different networks.

Table 2 shows the values of the costs (calculated as described above) for the three experiments. In every case, the β-continuation network produced the same cost and the same solution as the annealing algorithm. We ran many more experiments than are included here, but the β-continuation always found the same minimum as the annealing algorithm. Although this is not a proof of optimality for the β-continuation, it does indicate that if one is trying to deterministically minimize a cost function such as in (1), the β-continuation performs extremely well.

If the cost function were the last word on image smoothing and segmentation, we could immediately recommend the β-continuation. However, the qualitative results of the experiments say otherwise. We do not have sufficient space to show enough figures to effectively convey the different behaviors of the networks. However, observe Figures 11 and 12, which are

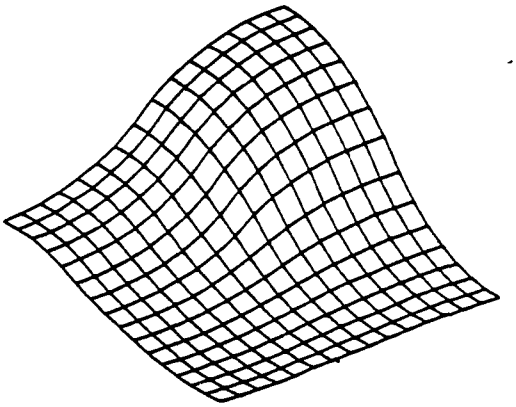

Figure 11: Network solution corresponding to global cost minimum in experiment 3.

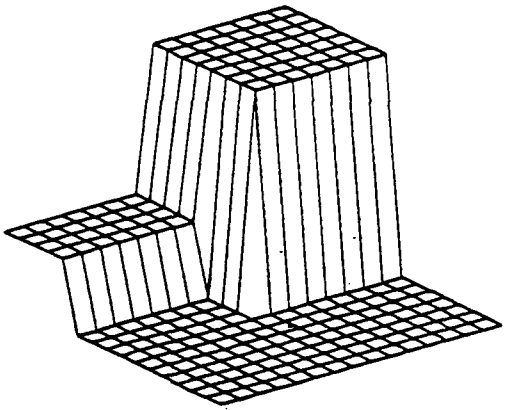

Figure 12: Network solution obtained by λ_f-continuation in experiment 3.

solutions obtained from Expt 3. The former is the global optimal solution, yet it is a completely smoothed "blob" which retains none of the edge information contained in the original image. The latter figure was produced by the λ_f-continuation. Although this solution does not correspond to the global cost minimum, in a visual sense it is the best solution since the original input image has been completely recovered.

Naturally, this calls into question the entire cost function methodology used for smoothing and segmentation. This difficulty arises because in trying to compute a minimizing solution to a given cost function, we are concentrating on only one portion of a larger problem. The total problem can be formulated in the following top-down manner:

1. What is the best form of cost function to use for a given vision task (in our case smoothing and segmentation)?
2. What is the best set of parameters for the cost function?
3. What is the best (minimizing) solution to the parameterized cost function?

Solving 3 without giving enough attention to 1 and 2 can produce useless results, as the third experiment demonstrates.

Some interesting properties of the λ_f-continuation network must be mentioned. For $\beta < \infty$, it can be shown that there exist a $\lambda_{min} > 0$ and a $\lambda_{max} < \infty$ such that for $\lambda_f > \lambda_{max}$ and

for $\lambda_f < \lambda_{min}$, the network has a unique solution. In fact, for $\lambda_f > \lambda_{max}$, the output will essentially match the input (i.e., $y \approx u$), whereas for $\lambda_f < \lambda_{min}$, the output will contain no edges. Consider the network behavior as a function of λ_f as λ_f is varied continuously from λ_{max} to λ_{min}. The initial solution of the network will closely match the input. Then, as λ_f is decreased, edges will begin to disappear, (first the smaller, then the larger), until all the edges are gone. In other words, λ_f acts as a scale-space parameter. This has important practical applications. The network of Perona and Malik has the property that *time* acts as a scale-space parameter. In contrast, we can exercise direct control over the scale-space parameter in the λ_f-continuation network. Finally, it turns out that when the edges disappear as λ_f is varied, they disappear in a catastrophic fashion. Investigation of the λ_f network in a bifurcation theory framework is a topic of ongoing research.

5 Conclusion

In this paper, we developed and compared a series of nonlinear networks for image smoothing and segmentation. The results of several experiments indicate that the typical cost (or energy) function minimization formulation of the smoothing and segmentation problem does not necessarily capture the essence of the task. For the specific parameter values we used, the λ_f-continuation network performed extremely well even though it did not always find the solution with minimum cost. The λ_f-continuation network has several implementation advantages over the β-continuation network. First, in certain cases, it seems to perform the smoothing and segmentation task in a more visually correct fashion. Second, λ_f can be used as a *scale-space parameter*. Finally, since the λ_f-continuation only requires that a linear resistance be varied, its VLSI implementation should be much more compact than that of the β-continuation (which would require that the characteristics of a nonlinear resistor be varied).

ACKNOWLEDGMENTS

This work was supported by the Defense Advanced Research Projects Agency under Contract No. N00014-87-K-825, the National Science Foundation under Grant No. MIP-88-14612, and E.I. DuPont de Nemours and Co. The first author was also supported by an AEA/Dynatech faculty development fellowship. The authors would like to acknowledge helpful discussions with Davi Geiger and Professor Jacob White. A result in [Geiger & Girosi] inspired the alternate derivation in Section 2.2.

References

[Blake & Zisserman] A. Blake and A. Zisserman, *Visual Reconstruction*, MIT Press, Cambridge, MA, 1987.

[Blake] A. Blake, "Comparison of the Efficiency of Deterministic and Stochastic Algorithms for Visual Reconstruction," *IEEE Trans. PAMI-11(1)*, pp. 2 - 12, January 1989.

[Cohen] F.S. Cohen and D.B. Cooper, "Simple Parallel Hierarchical and Relaxation Algorithms for Segmenting Noncausal Markovian Random Fields," *IEEE Trans. PAMI-9(2)*, pp. 195 - 219, March 1987.

[Elfadel] I. Elfadel, "Note on a Switching Network for Image Segmentation," unpublished manuscript, October 1988.

[Geman & Geman] S. Geman and D. Geman, "Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images," *IEEE Trans. PAMI-6(6)*, pp. 721 - 741, November 1984.

[Geiger & Girosi] D. Geiger and F. Girosi, "Parallel and Deterministic Algorithms from MRF's: Surface Reconstruction and Integration," MIT AI Laboratory Memo 1114, May, 1989.

[Harris *et al*] J. Harris, C. Koch, J. Luo, and J. Wyatt, "Resistive Fuses: Analog Hardware for Detecting Discontinuities in Early Vision," *Analog VLSI Implementation of Neural Systems*, C.A. Mead and M. Ismail, eds., Kluwer, 1989, pp. 27-56.

[Hillis] W.D. Hillis, *The Connection Machine*, MIT Press, Cambridge, MA, 1985.

[Horn] B.K.P. Horn, "Parallel Networks for Machine Vision," MIT AI Laboratory Memo 1071, August 1988.

[Koch *et al*] C. Koch, J. Marroquin, and A. Yuille, "Analog 'Neuronal' Networks in Early Vision," *Proc. Natl. Acad. Sci. USA*, vol. 83, pp. 4263-4267, 1986.

[Marroquin] J.L. Marroquin, "Optimal Bayesian Estimators for Image Segmentation and Surface Reconstruction," MIT AI Laboratory Memo 839, April 1985.

[Marroquin *et al*] J. Marroquin, S. Mitter, and T. Poggio, "Probabilistic Solution of Ill-Posed Problems in Computaional Vision," *Jour. Amer. Stat. Assoc. (Theory and Methods)*, vol. 82, no. 397, pp. 76-89, March 1987.

[Millar] W. Millar, "Some General Theorems for Non-Linear Systems Possessing Resistance," *Phil. Mag.*, 42:1150-1160, 1951.

[Ortega & Rheinboldt] J.M. Ortega and W.C. Rheinboldt, *Iterative Solution of Nonlinear Equations in Several Variables*, Academic Press, New York, 1970.

[Penfield *et al*] P. Penfield, Jr., R. Spence, and S. Duinker, *Tellegen's Theorem and Electrical Networks*, MIT Press, Cambridge, MA, 1970.

[Perona & Malik] P. Perona and J. Malik, "A Network for Multiscale Image Segmentation," *Proceedings of ISCAS '88*, pp. 2565-2568, 1988.

[Poggio & Koch] T. Poggio and C. Koch, "An Analog Model of Computation for the Ill-Posed Problems of Early Vision," MIT AI Laboratory Memo 783, May 1984.

[Tank & Hopfield] D.W. Tank and J. J. Hopfield, "Simple 'Neural' Optimization Networks: An A/D Converter, Signal Decision Circuit, and a Linear Programming Circuit," *IEEE Trans. CAS-33(5)*, May 1986.

[Terzopoulos] D. Terzopoulos, "Multigrid Relaxation Methods and the Analysis of Lightness, Shading, and Flow," MIT AI Laboratory Memo 803, October 1984.